ATTACHMENT A

SERK (Receptor Kinase SEQ ID NO: 21)

```
  1  messyvvfil lslillpnhs lwlasanleg dalhtlrvtl
 41  vdpnnvlqsw dptlvnpctw fhvtcnnens virvdlgnae
 81  lsghlvpelg vlknlqelys nnitgpipsn lgn[redacted]
121  [redacted]lsk lrflrlnnns ltgsipmslt
161  nittlqvldl snnrlsgsvp dngsfslftp isfannldlc
201  gpvtshpcpg sppfs[redacted]vstps gygitgaia[redacted]
241  [redacted]rkpldiff dvpaeedpev
281  hlgqlkrfsl relqvasdgf snknil[redacted]kvykgrla
321  dgtlv[redacted] keertpggel qfqtevemis mavhrnllrl
361  rgfcmtpter llvypymang svasclrerp psqppldwpt
401  rkrialgsar glsylhdhcd pkiihrdvka anilldeefe
441  avvg[redacted]ylstgks
481  se[redacted]
```

Secretion signal underlined

[redacted] Leucine rich region

[redacted] Proline box

[redacted] Transmembrane domain

[redacted] Protein kinase domain

[redacted] Subdomain I: Glycine triad

[redacted] Subdomain II: Invariant lysine

[redacted] Subdomain VIb: Catalytic loop

[redacted] Subdomain VII/VIII: Activation loop bounded by invariant DFG and APE motifs [redacted]

[redacted] Subdomain IX: Invariant d and g

ATTACHMENT B

Peptide Motifs and Protein Modules in Cell Signalling

A great leap in the understanding of cellular signal transduction pathways came with the realisation that...

- certain **linear amino acid sequences (or "motifs")**
- as well as certain **3-dimensional folded domains (or "modules")**

...are contained within the structures of (often unrelated) diverse proteins involved in signalling. Although a few of these motifs are found in proteins not involved in signalling, many are unique to signalling molecules.

Modules are tightly folded discrete structures, many of which can be inserted into unrelated proteins during evolution, without effect on the overall structure/function of the acceptor protein. SH-2 and SH-3 domains are examples of modules found in many unrelated types of proteins involved in signal transduction.

Searching protein databases for the presence of such motifs and modules allows identification of signalling functions in previously uncharacterised sequences.

1. Protein Kinases

Definitions...

Kinase:- an enzyme which catalyses the phosphorylation of an acceptor molecule, with ATP (usually) acting as the phosphate (phosphoryl) donor. You will be familiar with kinases in glycolysis which transfer phosphate to carbohydrates – e.g. hexokinase.

Protein kinases:- transfer phosphate to specific proteins. The phosphate either tags the protein or alters its subsequent activity.

There are basically two types of protein kinases.

(a). Serine/threonine protein kinases – which phosphorylate either serine or

threonine

(b). Tyrosine protein kinases – which phosphorylate tyrosines

Members of the tyrosine protein kinase family may be either receptor tyrosine kinases or non-receptor tyrosine kinases.

Both Ser/Thr- and Tyr kinases share a homologous stretch of approximately 300 amino acids which represents the core catalytic site

We shall use the insulin receptor as an example since it contains not only a tyrosine kinase domain, but also many other motifs and modules found in signal transduction molecules. The insulin receptor can be thought of as a dual-functional protein containing an extracellular recognition site for insulin binding and an intracellular catalytic site which phosphorylates tyrosines.

Figure 1.1. shows the complete human insulin receptor sequence. Note that numbering varies between papers depending on whether the signal sequence and/or the splice variant region are counted.

Figure 1.1. insulin receptor sequence

```
                  signal peptide  27|
      1 MGTGGRRGAA AAPLLVAVAA LLLGAAG

                  mature alpha-chain 1   4
                                     HLY PGEVCPGMDI RNNLTRLHEL ENCSVIEGHL
   34   61 QILLMFKTRP EDFRDLSFPK LIMITDYLLL FRVYGLESLK DLFPNLTVIR GSRLFFNYAL
   94  121 VIFEMVHLKE LGLYNLMNIT RGSVRIEKNN ELCYLATIDW SRILDSVEDN HIVLNKDDNE
  154  181 ECGDICPGTA KGKTNCPATV INGQFVERCW THSHCQKVCP TICKSHGCTA EGLCCHSECL
  214  241 GNCSQPDDPT KCVACRNFYL DGRCVETCPP PYYHFQDWRC VNFSFCQDLH HKCKNSRRQG
```

```
334  361 TVINGSLIIN IRGGNNLAAE LEANLGLIEE ISGYLKIRRS YALVSLSFFR KLRLIRGETL
394  421 EIGNYSFYAL DNQNLRQLWD WSKHNLTTTQ GKLFFHYNPK LCLSEIHKME EVSGTKGRQE
454  481 RNDIALKTNG DKASCENELL KFSYIRTSFD KILLRWEPYW PPDFRDLLGF MLFYKEAPYQ
514  541 NVTEFDGQDA CGSNSWTVVD IDPPLRSNDP KSQNHPGWLM RGLKPWTQYA IFVKTLVTFS
574  601 DERRTYGAKS DIIYVQTDAT NPSVPLDPIS VSNSSSQIIL KWKPPSDPNG NITHYLVFWE
634  661 RQAEDSELFE LDYCLKGLKL PSRTWSPPFE SEDSQKHNQS EYEDSAGECC SCPKTDSQIL
664  721 KELEESSFRK TFEDYLHNVV FVPRKTSSGT GAEDPRPSRK RR

                                                              736
                    mature beta-chain                          SLGDVGNV TVAVPTVAAF
754  781 PNTSSTSVPT SPEEHRPFEK VVNKESLVIS GLRHFTGYRI ELQACNQDTP EERCSVAAYV
814  841 SARTMPEAKA DDIVGPVTHE IFENNVVHLM WQEPKEPNGL IVLYEVSYRR YGDEELHLCV
874  901 SRKHFALERG CRLRGLSPGN YSVRIRATSL AGNGSWTEPT YFYVTDYLDV PSNIAKIIIG
934  961 FLIFVFLFSV VIGSIYLFLR KRQPDGPLGP LYASSNPEYL SASDVFPCSV YVPDEWEVSR
994 1021 EKITLLRELG QGSFGMVYEG NARDIIKGEA ETRVAVKTVN ESASLRERIE FLNEASVMKG
1054 1081 FTCHHVVRLL GVVSKGQPTL VVMELMAHGD LKSYLRSLRP EAENNPGRPP PTLQEMIQMA
1114 1141 AEIADGMAYL NAKKFVHRDL AARNCMVAHD FTVKIGDFGM TRDIYETDYY RKGGKGLLPV
1174 1201 RWMAPESLKD GVFTTSSDMW SFGVVLWEIT SLAEQPYQGL SNEQVLKFVM DGGYLDQPDN
1234 1261 CPERVTDLMR MCWQFNPKMR PTFLEIVNLL KDDLHPSFPE VSFFHSEENK APESEELEME
1294 1321 FEDMENVPLD RSSHCQREEA GGRDGGSSLG FKRSYEEHIP YTHMNGGKKN GRILTLPRSN
1354 1381 PS
```

grey = numbering of pro-form (before processing)

-27-0:-	Signal peptide (cleaved off during ER	1–27
1-735:-	Mature α-chain (ligand-binding, extracellular)	28–758
718-729:-	Splice variant region (missing in short isoform)	745–756
736-1355:-	Mature β-chain (catalytic and regulatory, cytosolic)	763-1382

Transmembrane domain (BOXED)

Catalytic domain

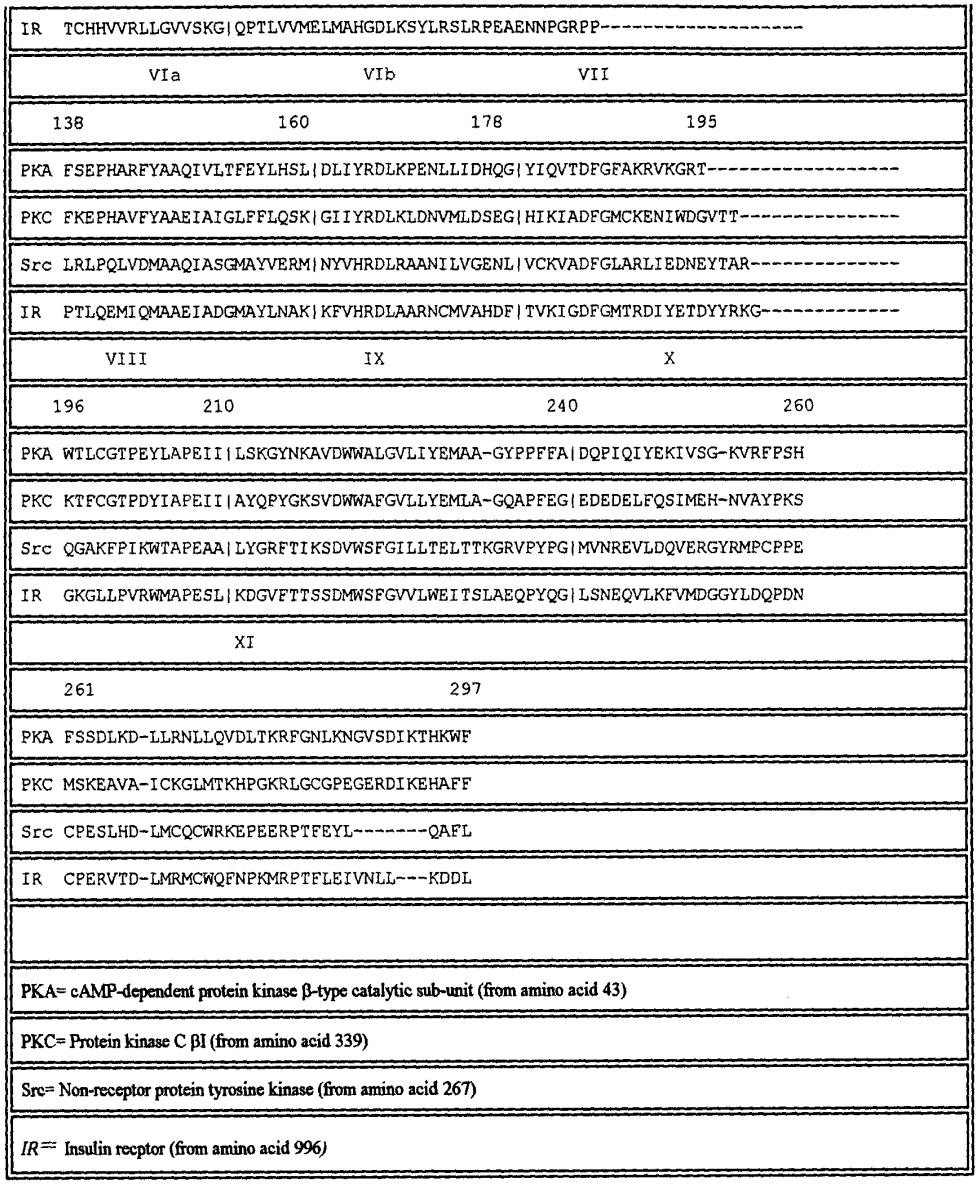
1003-1011:- Rossmann motif (tri-glycyl + lys, P-anchor)
1150-1179:- Activation segment (= activation- & P+1 loops)
1130-1139:- Catalytic loop

1.2. Structural and functional features shared by all protein kinase enzymes

The work of Steven Hanks led to the recognition that all protein kinases have conserved residues and homologous stretches centred on **12 sub-domains** within the approximately 300 amino acid kinase stretch (see Figure 1.2.)

Figure 1.2. ALIGNMENTS OF PROTEIN KINASE CATALYTIC SITES – SUB-DOMAIN ASSIGNMENTS ACCORDING TO Hanks (1988) Science

```
          I                                  II              III
  43                    64      65                  83          98
PKA FERKKTLGTGSFGRVMLVKHKA-----|TEQYYAMKILDKQKVVKLK|QIEHTLNEKRILQAV---------
PKC FNFLMVLGKGSFGKVMLSERKG-----|TDELYAVKILKKDVVIQDD|DVECTMVEKRVLALPG--------
Src LRLEVKLGQGCFGEVWMGTWNG-----|-TTRVAIKTLKPGTM----|SPEAFLQEAQVMKKL---------
IR  ITLLRELGQGSFGMVYEGNARDIIKGE|AETRVAVKTVNESASLR--|ERIEFLNEASVMKGF---------
       IV                  V
  99          113                     137
PKA NFPFLVRLEYAFKDN|SNLYMVMEYVPGGEMFSHLRRIGR----------------------------
PKC KPPFLTQLHSCFQTM|DRLYFVMEYVNGGDLMYHIQQVGR----------------------------
Src RHEKLVQLYAVVSE-|EPIYIVTEYMSKGSLLDFLKGETGKY--------------------------
```



See Steven Hanks Web site

1.3. Catalytic Domains of Protein Kinases

Not surprisingly, many of the conserved residues were found to have essential roles to play in catalysis. Of particular importance are three loops:- the 'P-loop' (sub-domain I); the 'C-loop' (sub-domain VIb) and the 'A-loop' (subdomains VII/VIII). See Figure 1.3.

Figure 1.3. Protein kinase catalytic site loops

	I	II	III
	(P-loop plus a lysine) = Rossmann Motif		
PKA	LGTGSFGRVMLVKHKA-----	TEQYYAMKILDKQKVVKLK	QIEHTLNEKRILQAV-
PKC	LGKGSFGKVMLSERKG-----	TDELYAVKILKKDVVIQDD	DVECTMVEKRVLALPG
Src	LGQGCFGEVWMGTWNG-----	-TTRVAIKTLKPGTM----	SPEAFLQEAQVMKKL-
InR	LGQGSFGMVYEGNARDIIKGE	AETRVAVKTVNESASLR--	ERIEFLNEASVMKGF-

	VIb	VII	VIII
	Catalytic loop	Activation segment (A-loop & P+1-loop)	
PKA	DLIYRDLKPENLLIDHQG	YIQVTDFGFAKRVKGRT-----	WTLCGTPEYLAPEII
PKC	GIIYRDLKLDNVMLDSEG	HIKIADFGMCKENIWDGVTT--	KTFCGTPDYIAPEII
Src	NYVHRDLRAANILVGENL	VCKVADFGLARLIEDNEYTAR-	QGAKFPIKWTAPEAA
InR	KFVHRDLAARNCMVAHDF	TVKIGDFGMTRDIYETDYYRKG	GKGLLPVRWMAPESL

1.3a. The Rossmann Motif
All kinases (including protein kinases as well as those which phosphorylate metabolites or lipids) contain a characteristic motif in their active site, called a "Rossmann Motif".
This consists of a triad of glycines:- **Gly**.Xxx.**Gly**.Xxx.Xxx.**Gly** *(Xxx=any amino acid)*, and a conserved lysine. [See:- Bossemeyer, D. (1994) *TIBS*, **19**: 201-205]
The Rossmann motif is also found in non-kinase proteins which bind mononucleotides (ATP,GTP) and dinucleotides (NAD,NADP,FAD).

• For example the guanine nucleotide-binding proteins (G-proteins) such as

Ras have Rossmann motifs in their nucleotide binding sites.

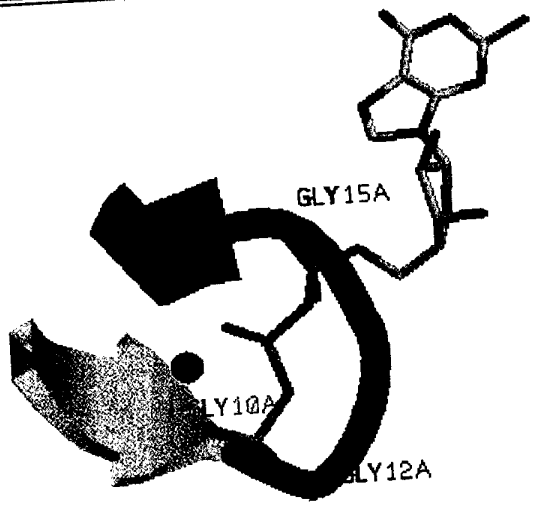
• In Harvey Ras, the sequence:-
Gly.Ala.**Gly**.Gly.Val.**Gly**.Lys.Ser is found

in Loop 1 between β strand 1 and the begining of α helix 1 (residues 10-17,

the "P-Loop"). See Figure 1.4. In kinases the loop is between two β-

strands

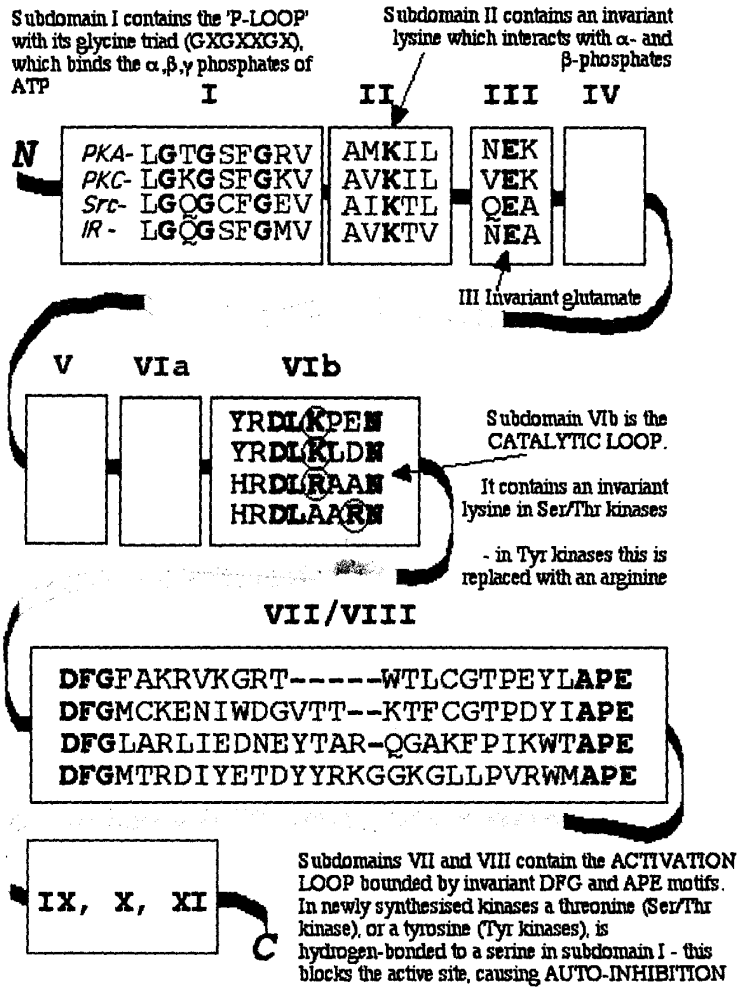
Figure 1.4. The Rossmann motif of Ras with GTP analogue bound



1.2b. P-loop, Catalytic loop and Activation Segment

These modules make up the functional active site. Subdomains important for catalytic function are shown in Figure 1.5.

Figure 1.5. Subdomain structure of protein kinase catalytic domains



Catalytic activity and auto-inhibition mechanisms

[Hanks, S.K., Quinn, A.M. & Hunter, T. (1988) *Science*, **241**: 42-52; Johnson, L.N., et al., (1996) *Cell*, **85**: 149-158; Frankel, M., et al. (1999) *Protein Science*, **8**: 2158-2165}]

A common feature of protein kinases is that they require a residue in the **ACTIVATION LOOP** to be phosphorylated before they can become activated.

(a)Some protein kinases are simply controlled by phosphorylation and de-phosphorylation of these activation loop residues – examples are MAP kinase and the insulin receptor tyrosine kinase.

(b) Other kinases, especially **those controlled by soluble second messengers** (e.g. PKA and PKC), are synthesised, then activated by autophosphorylation, whilst still being processed. The mature forms of PKC and PKA are phosphorylated on equivalent threonine residues (Thr197 in PKA) in their **activation loops**, but then become auto-inhibited by a different, secondary mechanism – the binding of **'PSEUDOSUBSTRATE SEQUENCES'** to their active sites (see later lectures).

Catalytic-activation loop interactions

- **Sub-domain I** consensus sequence: **Gly-X-Gly-X-X-Gly** (aa's:-50-55 in PKA) wraps around the phosphates of ATP, the amide bond nitrogens of the glycines providing a positively-charged electrostatic field which binds α and β phosphates. A serine **H-bonds** to either pseudosubstrate sequences or **autophosphorylation sites** (often found in subdomain VIII).

- **Sub-domain II** contains an **invariant Lys** (corresponds to Lys72 of PKA) which binds to α and β-phosphates. The lysine is held in position by a salt-bridge with Glu91.

• **Sub-domains VIb** represents the **CATALYTIC LOOP** (164-171 of PKA). Note that the **Lys** which co-ordinates the gamma phosphate in **Ser/Thr kinases** is replaced by **Arg in Tyr kinases.Lysine168 interaction with the** γ-phosphatestabilises the transition state. The invariant aspartate166 is theorised to be the catalytic residue. It acts as a base to remove a proton from the hydroxyl group of either serine/threonine or tyrosyl residues of the protein substrate, leaving an alcoholate or phenolate ion to particiapate in nucleophilic attack on the γ-phosphate of ATP.

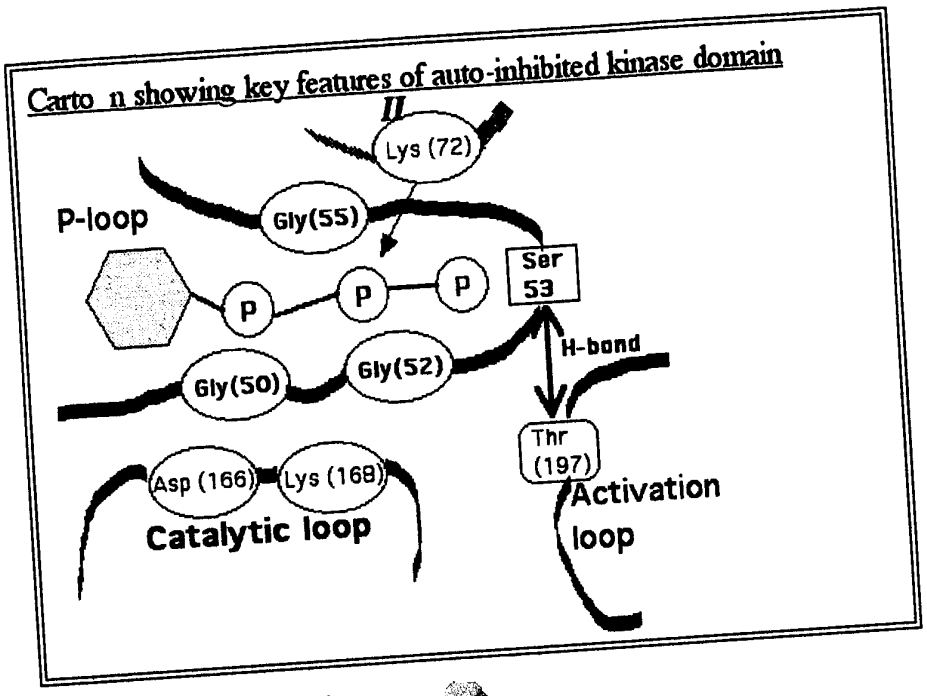
• **Sub-domain VIII** represents the **ACTIVATION LOOP** (184-208 of PKA). It contains consensus triplet: Ala-Pro-Glu..(**A.P.E**) its deletion in Src leads to an inactive kinase. Residues in this subdomain are often autophosphorylated as part of the activation mechanism (*See insulin receptor*).

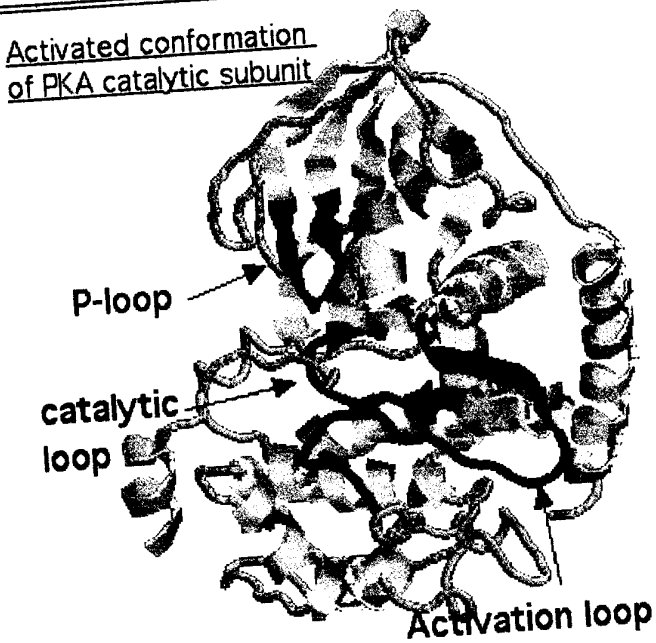
Activation loop blocks active site in un-phosphorylated form

- Threonine197 of Protein Kinase A (**PKA**) is H-bonded to the serine 53 of the P-loop. This blocks the binding of PKA's protein substrates.

- PKA autophosphorylates the threonine and the now negatively charged phosphothreonine is ejected from the active site and binds instead to arginine165.

- The activation loop has swung out of the active site and the kinase can now accomodate its normal protein substrates in its active site.


Carto n showing key features of auto-inhibited kinase domain
Lys (72)
P-loop
Gly(55)
Ser 53
P
P
P
H-bond
Gly(50)
Gly(52)
Thr (197)
Asp (166)
Lys (168)
Catalytic loop
Activation loop


Activated conformation of PKA catalytic subunit
P-loop
catalytic loop
Activation loop

ATTACHMENT C

Molecule: SeqID20.txt, 4081 bps DNA
Description:
File Name: seq id 20.cm5, dated 16 Oct 2001
Printed: 1-4081 bps (Full), format Annotated: Enzymes, Genes

XcmI

1 tctagaaacc ttttgatcat aatgaaaata aagagtccat ccaccacatg

51 gggtaagcat aatgtgtgat atttaaaggg taacaaatgt aatctgcttt

101 ttattttact ttttacctct actcaaattg tatgggcagt ttttttttt

151 ttttaaatga taagacaagt atctgtttaa tggtattgtg atgaaacagt

BseSI

201 agtaaagtca tatcgggcac gccatactac ttccacagtg gaacttggcc

BsmBI HincII

251 aaattttgtc tttgccgtct ctacagtttc ttccaccaaa ttttttgttg

301 acaaaactca aatctttcaa tctcatctct gccaaagttg ggtttagaaa

351 gaatatcagc aaacactaat atctttattg ttgcatggtt tatcaatcac

401 aaaattcaca accattgtaa aaaaaaattc acatttttgg tatgagattg

451 ctcacatgat agtgaacctc tttaacattt taactttact ttcataaata

501 cgggattacg aatcttactt gcattaaaaa tttagaaaag gtttttctac

SanDI
PpuMI
Eco0109I

551 ttaaagaaaa aagggaccca acagagagag gtttgaccag gagaaacggg

601 tgcatagcct taagagcttt caactacttt accccaaacc caaagcgatg

AgeI

651 tcactttcaa ccatctcttc tctcccccga acccgttttt ttgaccggtc

BbsI

701 agttcgggca gcagcaccgt tacgggcagc ttatattcct cgtcttcctc

SphI

751 ctctacacca ctgcatgccc ataaataaag cccgttgaga tctttaaaaa

801 tattaaataa tatatcaacg aaaaagctat tttattcata agaagaaaaa

851 gagaggaaca acaacaacac actaatcata gtttctctgg caggcttgtt

901 gttgcggctt aataaaaagc tcttttgtta ttattacttc acgtagattt

951 tccccaaaaa gctcttattt ttttgtttaa aaaaaaaagt ttcatcttta

1001 ttcaactttt gttttacagt gtgtgtgtga gagagagagt gtggtttgat

1051 tgaggaaaga cgacgacgag aacgccggag aattaggatt tttattttat

1101 tttttactct ttgtttgttt taatgctaat gggtttttaa aagggttatc

1151 gaaaaaatga gtgagtttgt gttgaggttg tctctgtaaa gtgttaatgg

1201 tggtgatttt cggaagttag ggttttctcg gatctgaaga gatcaaatca

1251 agattcgaaa tttagcattg ttgtttgaaa tggagtcgag ttatgtggtg

»».....exon 1.....»

1301 tttatcttac tttcactgat cttacttccg aatcattcac tgtggcttgc

».....exon 1.....»

1351 ttctgctaat ttggaaggtt cgtggttact caattactca gctttactcg

».....exon 1.....»»

1401 tttctcaatt actttctcga ttctttttta tttggaggtg aatcgctatc

1451 tttagtgtct gcattttgat ttatgaaaat tgttgttgtt ctttgtattt

TatI

ScaI

1501 gtaagattta gtggctagta ctttgaatac actgttttgc ttttcttgtt

1551 cagatcaact ttgtatattg taaaggcatg ttctttgggt tgaaaagctg

EcoRV

1601 ggttatttga tatcttaaga ttgatgttgt tgatccaaac attctctgaa

1651 agacttcatt tgtttttggt tttgtaaaga atttgtttaa ttattagcct

StuI

1701 ctaatctcag agaggcctgt ttgaatagtt ctctcttgaa attagacttt

MunI

1751 tcaccaattg atgctaattg tgtagatttg ttgttcttgt tataggtgat

1801 gctttgcata ctttgagggt tactctagtt gatccaaaca atgtcttgca
» exon 2 »

BamHI

1851 gagctgggat cctacgctag tgaatccttg cacatggttc catgtcactt
» exon 2 »

1901 gcaacaacga gaacagtgtc ataagagtgt aaagctttct tctactaatc
» exon 2 »»

Bpu10I
BbvCI BstEII

1951 ccacttttta aactttgacc tcagcgtggt taccgacatt tttgtttctt

BsmI

2001 ttgtcaaata cagtgatttg gggaatgcag agttatctgg ccatttagtt
»» exon 3 »

Ppu10I

2051 ccagagcttg gtgtgctcaa gaatttgcag tatttgtaag ttccacttat
» exon 3 »»

NsiI
BfrBI

2101 gcatcatgct ttaacaaaac aaatccaaga tttgacagaa gaagcactgg

2151 agttaccttt tgtaattgaa atcttttaa caagtttctt attttcttac

2201 agggagcttt acagtaacaa cataactggc ccgattccta gtaatcttgg
»» exon 4 »

2251 aaatctgaca aacttagtga gtttggatct ttacttaaac agcttctccg
» exon 4 »

AccIII

2301 gtcctattcc ggaatcattg ggaaagcttt caaagctgag atttctgtga

»....exon 4....»»

NdeI

Bst1107I

2351 gtatacatat gctttaccgg ctcagttaca gtctttgttt aatcttaggt

NgoMIV

2401 tttgttccaa tttttgactc tttgctgaaa attttacatg caagaatagc

»

NaeI

2451 cggcttaaca acaacagtct cactgggtca attcctatgt cactgaccaa

»....exon 5....»

2501 tattactacc cttcaagtgt tgtgagtcct ctcattaact ttcatttatg

»....exon 5....»»

2551 tctacttcat tctccctcag ttgatttgtt gagttaatgc acttaacctt

2601 gatggatgca acacagagat ctatcaaata acagactctc tggttcagtt

»»....exon 6....»

2651 cctgacaatg gctccttctc actcttcaca cccatcaggt tctatgattt

»....exon 6....»»

2701 atcctcttca gttatttcag ttgttgtgtc agtgtctgaa cttattctga

BsgI

2751 aactttcatt tccttgtgca gttttgctaa taacttagac ctatgtggac

»»....exon 7....»

2801 ctgttacaag tcacccatgt cctggatctc ccccgttttc tcctccacca

»....exon 7....»

2851 ccttttattc aacctccccc agtttccacc ccgagtaagc ctcctctttt

»....exon 7....»»

2901 tagtttacat tataggaaac agaagatgaa atctttgctt ctctgtcaat

BglI

2951 ccttttttctc atataactca tcttgccaat aaggcaataa ccaaatgatc

3001 taatttgatt tcaggtgggt atggtataac tggagcaata gctggtggag
»».....exon 8.....»

BstAPI
BspMI
AarI
PstI
3051 ttgctgcagg tgctgctttg ctctttgctg ctcctgcaat agcctttgct
».....exon 8.....»

3101 tggtggcgac gaagaaagcc actagatatt ttcttcgatg tgcctggtga
».....exon 8.....»»

3151 gtttattatt cgcattagtt tctgttctta gccagcaatt ttgttttgca

3201 gaaaagtatt ggaacaactg ttaatgaaaa tcaatacata agtcattgtt

3251 ttttaagtta caaactcttt tgagtaaaat ctcgattgca aaatctctat

3301 gcagccgaag aagatccaga agttcatctg ggacagctca agaggttttc
»»....exon 9.....»

3351 tttgcgggag ctacaagtgg cgagtgatgg gtttagtaac aagaacattt
».....exon 9.....»

3401 tgggcagagg tgggtttggg aaagtctaca agggacgctt ggcagacgga
».....exon 9.....»

SapI
3451 actcttgttg ctgtcaagag actgaaggaa gagcgaactc caggtggaga
».....exon 9.....»

SacI
Ecl136II
BanII
3501 gctccagttt caaacagaag tagagatgat aagtatggca gttcatcgaa
».....exon 9.....»

BsiEI
3551 acctgttgag attacgaggt ttctgtatga caccgaccga gagattgctt
».....exon 9.....»

3601 gtgtatcctt acatggccaa tggaagtgtt gcttcgtgtc tcagaggtaa
».....exon 9.....»»

BtrI

3651 aaactaaaca attaaacatc ttgtgctctc tctcaattac tttgacgtga

3701 agtgtttttt catgttttcc tttatgggtt cataattgtt ggttacacta

3751 atgacacaga gaggccaccg tcacaacctc cgcttgattg gccaacgcgg

»»»....exon 10....»

XhoI

Bpu1102I

3801 aagagaatcg cgctaggctc agctcgaggt ttgtcttacc tacatgatca

»....exon 10....»

3851 ctgcgatccg aagatcattc accgtgacgt aaaagcagca aacatcctct

»....exon 10....»

EcoRI

3901 tagacgaaga attcgaagcg gttgttggag atttcgggtt ggcaaagcta

»....exon 10....»

PmlI

OliI

BanI

BtgI

3951 atggactata aagacactca cgtgacaaca gcagtccgtg gcaccatcgg

»....exon 10....»

4001 tcacatcgct ccagaatatc tctcaaccgg aaaatcttca gagaaaaccg

»....exon 10....»

4051 acgttttcgg atacggaatc atgcttctag a

»....exon 10....»»